PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/29, 15/62, 15/82 C12P 21/02, C12N 5/10 A01H 5/00	A1	(11) International Publication Number: WO 93/21320 (43) International Publication Date: 28 October 1993 (28.10.93)

(21) International Application Number: PCT/CA92/00161

(22) International Filing Date: 15 April 1992 (15.04.92)

(60) Parent Application or Grant
(63) Related by Continuation
US 659,835 (CIP)
Filed on 22 February 1991 (22.02.91)

(71) Applicant *(for all designated States except US)*: UNIVERSITY TECHNOLOGIES INTERNATIONAL, INC. [CA/CA]; ES620, 2500 University Drive, N.W., Calgary, Alberta T2N 1N4 (CA).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)* : MOLONEY, Maurice, M. [GB/CA]; 131 Edgehill Place, Calgary, Alberta T3A 2S4 (CA).

(74) Agent: CLARK, Geoffrey, C.; Fetherstonhaugh & Co., 456 - 409 Granville Street, Vancouver, British Columbia V6C 1H5 (CA).

(81) Designated States: AT, AU, BB, BG, BR, CA, CH, CS, DE, DK, ES, FI, GB, HU, JP, KP, KR, LK, LU, MG, MN, MW, NL, NO, PL, RO, RU, SD, SE, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LU, MC, NL, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
With international search report.

(54) Title: OIL-BODY PROTEINS AS CARRIERS OF HIGH-VALUE PEPTIDES IN PLANTS

(57) Abstract

Methods and compositions for use therein are described for expressing a polypeptide of interest in a seed cell as a fusion protein with an oil body protein. By this means, the fusion protein is targeted to the oil bodies of a seed cell. The oil body is easily separated from other cellular material following lysis of the seed cell, for example by using the partitioning/surface properties of the oil body. The fusion protein may be isolated for example by affinity chromatography using antibodies directed to the oil body protein. Where desired, the polypeptide of interest can be recovered by treatment of the fusion protein with for example a protease capable of recognizing a proteolytic recognition site in the oil body protein proximal to the N-terminus of the polypeptide of interest.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SK	Slovak Republic
CI	Côte d'Ivoire	KZ	Kazakhstan	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslovakia	LK	Sri Lanka	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	UA	Ukraine
DK	Denmark	MG	Madagascar	US	United States of America
ES	Spain	ML	Mali	VN	Viet Nam
FI	Finland	MN	Mongolia		

OIL-BODY PROTEINS AS CARRIERS OF HIGH-VALUE PEPTIDES IN PLANTS

INTRODUCTION

Technical Field

This invention relates to a method for the production by recombinant means of a protein of interest which is easily purified from host cell components. The method is exemplified by expression of the protein of interest in plants, particularly seeds, as a chimeric peptide comprising an oil-body protein and the protein of interest.

Background

A variety of proteins have been expressed in plants. However, while the general feasibility of obtaining expression of foreign proteins in plants has been demonstrated, obtaining purified proteins from this source has some limitations. These limitations include the purification step necessary to obtain pure protein essentially free of plant derived materials and the degradation that may occur in extracts prepared during the purification procedure when the recombinant proteins obtained are in contact with aqueous buffers.

Plants bearing oilseeds such as soybean, rapeseed, sunflower and a number of other plant species such as corn, carrot, etc., store triglycerides in their seeds. In the plant, these triglycerides act as a source of energy for a germinating seed and the subsequent seedling. The triglycerides are widely used as vegetable oils in foods and in food preparation and also for some industrial applications.

Triglycerides are immiscible with water and partition by floating on the surface of aqueous solutions or by forming small globules or liposomes as a suspension in

the aqueous phase. Such globules will naturally coalesce if they are not stabilized by a modified surface layer. This coalescence can result in a suspension of globules of random sizes. In seeds, when triglyceride is stored, the oil globules are actually encapsulated lipid or oil bodies normally of uniform size. Associated with the surface of these oil bodies is a half unit membrane studded with several proteins, generally referred to as oil-body proteins.

At least one class of oil-body proteins has some characteristics which are highly conserved between species. This class of oil-body proteins is referred to as an "oleosin." The hydrophilic N- and C-termini of these proteins appear to be quite divergent, whereas the lipophilic internal region (central core) appears to be highly conserved between species. The oleosins are strongly associated with the oil bodies; this strong association to the oil-bodies may, in major part, be due to the lipophilic nature of these central core. It is therefore of interest to determine whether oil body proteins such as oleosins may be useful in a method for the production of recombinant proteins by providing a means for separation of the recombinant proteins from plant derived materials.

Relevant Literature

The production of foreign (recombinant) peptides in plants has been investigated using a variety of approaches including transcriptional fusions using a strong constitutive plant promoter (e.g., from cauliflower mosaic virus--Sijmons *et al.* (1990) *Bio/Technology*, *8*:217-221) and the coding of a foreign protein; transcriptional fusions with organ specific sequences (Radke *et al.* (1988) *Theoret. Appl. Genet.*, *75*:685-694); and translational fusions which require subsequent cleavage of a recombinant protein (Vander Kerkove *et al.* (1989) *Bio/Technology*, 7:929-932). Foreign proteins which have been expressed in plant cells include active proteins from bacteria (Fraley *et al.* (1983) *Proc.*

Nat'l. Acad. Sci. USA, *80*:4803-4807), animals (Misra and Gedamu (1989) *Theor. Appl. Genet.*, *78*:161-168), fungi and other plant species (Fraley *et al.* (1983) *Proc. Nat'l. Acad. Sci. USA*, *80*:4803-4807).

Some proteins, normally markers of integration, have been expressed in a tissue-specific manner, including some in seeds (Sen Gupta-Gopalan *et al.* (1985) *Proc. Nat'l. Acad. Sci. USA*, *82*:3320-3324); Radke *et al.* (1988) *Theor. Appl. Genet.*, *75*:685-694). These reports have concentrated specifically on the use of seed-storage protein promoters as a means of deriving seed-specific expression. Using such a system, Vanderkerkove *et al.* (1989) *Bio/Technol.*, *7*:929-932, expressed a high value peptide (leu-enkephalin) in seeds of *Arabidopsis thaliana* and *Brassica napus*. The yield of this peptide was quite low, but demonstrates the feasibility of expression of an animal peptide hormone in plant tissues. Maize oleosin has been expressed in seed oil bodies in *Brassica napus* transformed with a maize oleosin gene. The gene was expressed under the control of regulatory elements from a *Brassica* gene encoding napin, a major seed storage protein. The temporal regulation and tissue specificity of expression was reported to be correct for a napin gene promotor/terminator. See, Lee *et al.*, *Proc. Nat'l. Acad. Sci.* (USA) (1991) 88:6181-6185.

The oil globules which are produced in seeds all appear to be of a similar size, indicating that they are stabilized (Huang A.H.C. (1985) *in Modern Meths. Plant Analysis*, Vol. *1*:145-151 Springer-Verlag, Berlin). On closer inspection, it has been found that these are not simple oil-globules, but rather oil-bodies surrounded by a membrane. These oil-bodies have been variously named by electron microscopists, oleosomes, lipid bodies and spherosomes (Gurr MI. (1980) *in The Biochemistry of Plants*, *4*:205-248, Acad. Press, Orlando, Fla). The oil-bodies of a few species have been studied and the general conclusion is that they are encapsulated by an unusual "half-unit" membrane comprising not a classical lipid bilayer, but rather a single amphophilic layer with hydrophobic groups r

the inside and hydrophilic groups on the outside (Huang A.H.C. (1985) *in Modern Meths. Plant Analysis*, Vol. *1*:145-151 Springer-Verlag, Berlin).

Analysis of the contents of lipid bodies has demonstrated that apart from triglyceride and membranous material, there are also several polypeptides/proteins associated with the surface or lumen of the oil body (Bowman-Vance and Huang (1987) *J. Biol. Chem.*, *262*:11275-11279, Murphy *et al.* (1989) *Biochem. J.*, *258*:285-293, Taylor *et al.* (1990) *Planta*, *181*:18-26). Oil-body proteins have been identified in a wide range of taxonomically diverse species (Moreau *et al.* (1980) *Plant Physiol.*, *65*:1176-1180; Qu *et al.* (1986) *Biochem. J.*, *235*:57-65) and been shown to be uniquely localized in oil-bodies and not found in organelles of vegetative tissues. In *Brassica napus* (rapeseed) there are at least three polypeptides associated with the oil-bodies of developing seeds (Taylor *et al.* (1990), *Planta*, *181*:18-26). The numbers and sizes of oil-body associated proteins may vary from species to species. In corn, for example, there are two immunologically distinct polypeptide classes found in oil-bodies (Bowman-Vance and Huang (1988) *J. Biol. Chem.*, *263*:1476-1481). Oleosins have been shown to comprise regions of alternate hydrophilicity, hydrophobicity and hydrophilicity (Bowman-Vance and Huang (1987) *J. Biol. Chem.*, *262*:11275-11279). The amino acid sequences of oleosins from corn, rapeseed, and carrot have been obtained. See Qu and Huang (1990) *J. Biol. Chem.*, *265*:2238-2243, Hatzopoulos *et al.* (1990) *Plant Cell*, *2*:457-467, respectively. In an oilseed such as rapeseed, oleosin may comprise between 8% (Taylor et al. (1990) *Planta*, *181*:18-26) and 20% (Murphy *et al.* (1989) *Biochem. J.*, *258*:285-293) of total seed protein. Such a level is comparable to that found for many seed storage proteins.

Genes encoding oil-body proteins have been reported for two species, maize (*Zea mays*, Bowman-Vance and Huang (1987) *J. Biol. Chem.*, *262*:11275-11279; and Qu and

Huang (1990) *J. Biol. Chem.*, *265*:2238-2243) and carrot (Hatzopoulos *et al.* (1990) *Plant Cell*, *2*:457-467).

SUMMARY OF THE INVENTION

Methods and compositions are provided for the production of peptides which may be easily purified from host proteins. The method includes the steps of preparing a chimeric DNA construct which includes a sequence encoding an oil-body specific sequence comprising the coding sequence of a seed-specific oil-body protein gene, or a sequence encoding at least a portion of the hydrophobic core of an oil-body protein, and a coding sequence for a peptide of interest from which an expression cassette containing the chimeric DNA construct can be prepared; transforming a host cell with the expression cassette under genomic integration conditions; and growing the resulting transgenic plant to produce seed in which the polypeptide of interest is expressed as a fusion protein with the oleosin.

The polypeptide of interest may be purified by isolating oil-bodies from the cells of the seed, and disrupting the oil-bodies so that the fusion protein is released. The oil-body protein is then easily separated from other proteins and plant derived material by phase separation. Optionally a cleavage site may be located at least one of prior to the N-terminus and after the C-terminus of the polypeptide of interest allowing the fusion polypeptide to be cleaved and separated by phase separation into its component peptides. The production system thus provides for targeting of the chimeric peptide by its oil-body protein functionality to the oil bodies which, in turn, permits rapid purification of the polypeptide of interest. This production system finds utility in the production of many peptides such as those with pharmaceutical, enzymic, rheological and adhesive properties.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1A. shows the nucleotide sequence and deduced amino-acid sequence (17 kDa protein) of an oil-body protein gene (oleosin) from *Arabidopsis thaliana*. Underlined are the direct repeats (R1 and R2) and inverted repeat (T), a CACA, TATA, TAAT and polyadenylation signals. The intron sequence is printed in lower case and a putative ABA-binding site is indicated in bold.

Fig. 1B. shows a comparison of the sequences of oil-body 16 Kd protein from carrot, an 18 Kd and 16 Kd oil-body protein from maize and a 17 Kd oil-body protein from *Arabidopsis thaliana* indicating conserved and divergent regions of the proteins; the amino acid sequences are aligned to show the conservation of sequence in the central region of the proteins.

Fig. 2. shows constructs used for the fusion of oil-body protein genes with genes encoding foreign peptides. IA is a C-terminal fusion of a desired peptide to OBP; IB is an N-terminal fusion of a desired peptide to OBP; II is an internal fusion of a desired peptide within OBP; and III is an inter-dimer translational fusion of desired peptide enclosed between two substantially complete oil body protein targeting sequences. In the upper portion of Figure (A) are shown the DNA constructs used for translational fusions of desired peptides to oil-body proteins. In the lower portion of Figure (B) are shown the configurations of the gene products, shown on the upper portion of the translation and the delivery to the oil bodies. The key to the figure is as follows: bottom left-top right hatched box represents an OBP promoter or other seed specific promoter; bottom right-top left hatched box represents a desired peptide coding sequence; open box represents an oil-body protein coding sequence or synthetic targeting sequence based on OBP conserved motifs; vertical-horizontal hatched box represents a gene terminator containing a polyadenylation signal; hatched circle

represents a protease recognition motif; corkscrew line represents a native C- or N-terminal of OBP.

Fig. 3. shows a detailed arrangement for construction of a C-terminal fusion. Shown is the arrangement is a collagenase recognition motif coding sequence as a linker in the fusion of a typical oil-body protein gene and a fusion peptide, to be linked here using an *NcoI*, for cloning and expression in plants.

Fig. 4. shows schematically the process of construction of fusion peptide vectors, their introduction into plants and subsequent extraction and assay of the desired recombinant peptide.

Figure 5 shows a schematic representation of the construction of pCGOBPILT. The broken line box represents an oleosin promoter; the top left-bottom right hatched box represents an oleosin coding sequence; the horizontal-vertical hatched box represents an intron; the dotted box represents a 3' non-translated sequence; and the widely-spaced top left-bottom right hatched box represents an interleukin-1-β sequence equipped with a sequence encoding a protease cleavage site (Factor Xa or thrombin immediately upstream).

Figure 6 shows the design of oligonucleotide GVR11. In Figure 3A represents the 3' coding sequence of the *A. thaliana oleosin*, translationally fused to the factor Xa/IL-1-β coding sequence followed by a TAA stop codon. For future cloning purposes, a *PvuI* and *SalI* restriction enzyme recognition site are included. The creation of a *PvuI* restriction site resulted in the additional coding sequence for an alanine (ala). Underlined are the restriction enzyme recognition sequences. Overlined are the *A. thaliana* oleosin sequences and the factor Xa recognition sequence. The actual cleavage site is indicated with an asterisk (*). In Figure 3B, the sequence of GVR11 is shown. In order to make a fusion with the *A. thaliana* oleosin, the primer GVR11 needs to be a sequence complementary to the top strand.

Figure 7 shows the nucleotide sequence of OBPILT. Underlined is the sequence encoding IL-1-β; the sequence encoding the factor Xa recognition site is indicated in bold. The nopaline synthase terminator sequence is indicated in lower case letters.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

In accordance with the subject invention, methods and compositions are provided for production of peptides which are easily purified. The subject method includes the steps of preparing an expression cassette containing DNA sequences encoding a sufficient portion of an oil body specific sequence, such as an oleosin, to provide for targeting to an oil body and the peptide of interest; transforming the expression cassette into a plant cell host; generating a transgenic plant and growing it to produce seed in which the chimeric protein is expressed and translocated to the oil bodies. The chimeric peptide comprises the peptide of interest and an oil body protein such as an oleosin. The peptide of interest generally is a foreign peptide normally not expressed in seeds or found on the oil-body. The use of an oil-body protein as a carrier or targeting means provides a simple mechanism to obtain purification of the foreign protein. The chimeric protein is separated away from the bulk of cellular protein in a single step (such as centrifugation or flotation); the protein is also protected from degradation during extraction as the separation also removes non-specific proteases from contact with the oil-bodies. The gene encoding the foreign peptide may be derived from any source, including plant, bacterial, fungal or animal source. Desirably, the chimeric peptide will contain sequences which allow for cleavage of the peptide of interest from the oleosin. The method may be employed to express a variety of peptides which are then easily purified.

Targeting a foreign, recombinant protein to the oil-body imparts several advantages, including the following. The protein can be separated from the bulk of cellular contents after cell lysis by centrifugation. The oil-body fraction will float on the surface of the extract. The protein can optionally be provided with a peptide linker containing a protease recognition site. This permits release of the peptide from the oil-body. The protein can be introduced into a recombinant polypeptide in such a way that it is within a lipophilic conserved region. This results in the internalization of the recombinant peptide into the oil-body, thus protecting it from protease attack.

The expression cassette generally will include in the 5'-3' direction of transcription, a transcriptional and translational regulatory region capable of expression in developing seed, typified by the promoter and upstream regions associated with an oil body protein, which will provide for expression of the chimeric protein in seed, a DNA sequence encoding a chimeric peptide comprising an amino acid sequence to provide an oil body targeting means and a protein of interest, and a transcriptional and translational termination region functional in plants. One or more introns may also be present.

The oil-body specific sequence finds analogy in fragments of oil-body proteins, particularly oleosins. The oil-body specific sequence may be the same as that of a sequence obtainable from an oil-body protein, are which has sufficient homology to provide for the desired targeting of a protein of interest to an oil body. By "obtainable" is intended an amino acid sequence which may be natural, synthetic or a combination, sufficiently similar to a native oil body protein amino acid sequence to provide the desired targeting. Of particular interest is the central hydrophobic domain of oil body proteins which appears to be highly conserved among different plant species, and fragments thereof and homologous sequences at the amino acid level.

The deduced amino acid sequence for an *Arabidopsis thaliana* oil-body protein is as follows:

```
                                10                  20
M-N-G-R-D-R-D-Q-Y-Q-M-S-G-R-G-S-D-Y-S-K-
                                30                  40
S-R-Q-I-A-K-A-A-T-A-V-T-A-G-G-S-L-L-V-L-
                                50                  60
L-S-L-T-L-V-G-T-V-I-A-L-T-V-A-T-P-L-L-V-
                                70                  80
I-S-S-T-I-L-V-P-A-L-I-T-V-A-L-L-I-T-G-S-
                                90                  100
L-S-S-G-G-F-G-I-A-A-I-T-V-F-S-W-I-Y-K*Y-
                                110                 120
L-L-I-E-H-P-Q-G-S-D-K-L-D-S-A-R-M-K-L-G-
                                130                 140
S-K-A-Q-D-L-K-D-R-A-Q-Y-Y-G-Q-Q-H-T-G-W-
                                150
E-H-D-R-D-R-T-R-G-G-Q-H-T-T
```

Amino acids from about 25-101 comprise the central hydrophobic domain.

Of particular interest as a targeting means for some applications are oil-body specific sequences or fragments thereof of the following formula which provide for targeting to an oil body:

$$pp^{1}- aa^{25}- aa^{26}- \begin{matrix}V\\A\end{matrix} - \begin{matrix}V-\\A\end{matrix} T - \begin{matrix}L-\\A\end{matrix} aa^{21}- \begin{matrix}P-\\T\end{matrix}$$

$$aa^{34}- G- G- aa^{36} - \begin{matrix}L-\\M\end{matrix} L- aa^{39}- L- aa^{41}-$$

$$\begin{matrix}G\\S\end{matrix} - \begin{matrix}I\\L\end{matrix} - aa^{44}- L- aa^{46}- aa^{47}- \begin{matrix}T\\S\end{matrix} - \begin{matrix}L\\V\end{matrix} - \begin{matrix}I\\V\end{matrix} -$$

$$aa^{51}- L- aa^{53}- V- A- T- P- \begin{matrix}L\\V\end{matrix} - aa^{59} - \begin{matrix}L\\V\end{matrix} -$$

$$\begin{matrix}L\\I\end{matrix} - F- S- P- \begin{matrix}V\\I\end{matrix} - L- V- P- A- \begin{matrix}A\\L\end{matrix} - \begin{matrix}L\\I\end{matrix} - aa^{73} -$$

$$aa^{74} - aa^{75} - aa^{76} - aa^{77} - aa^{78} - G- F- L-$$

$$\begin{matrix} G \\ S \\ T \end{matrix} - S- aa^{87} - G- \begin{matrix} L \\ V \\ I \end{matrix} - aa^{89} - aa^{90} - \begin{matrix} L \\ I \end{matrix} - \begin{matrix} S \\ T \end{matrix} -$$

$$aa^{93} \; aa^{94} - \begin{matrix} S \\ T \end{matrix} - aa^{96} - aa^{97} - aa^{98} - aa^{99} -$$

$$aa^{100} - aa^{101} - pp^{2}$$

wherein:

PP^1 and PP^2 are the same or different, and may be the same as or different from a natural oil-body protein, usually different; they may be hydrogens, indicating the terminal portion of the indicated polypeptide or may be polypeptides having a total of up to 1000 amino acids, more usually of up to about 500 amino acids, and may have a total of as few as 1 amino acid, or may individually or separately be polypeptides of from 1-100 amino acids, more usually from about 1-75 amino acids, more particularly from about 5-50 amino acids; these polypeptides will have specific applications in modifying a specifically described sequence for a predetermined purpose;

aa^{25} may be any amino acid, particularly a neutral aliphatic amino acid, generally of 3-6 carbon atoms, more particularly leucine or alanine;

aa^{26} is a neutral aliphatic amino acid, particularly alanine or an hydroxy substituted amino acid of from 3-4 carbon atoms, particularly threonine or a basic amino acid of from 5-6 carbon atoms, particularly lysine;

aa^{31} is a neutral unsubstituted aliphatic amino acid of from 3-6 carbon atoms, particularly alanine, valine or leucine or an aromatic unsubstituted amino acid, particularly phenylalanine;

aa^{33} is a neutral unsubstituted aliphatic amino acid of from 3-6 carbon atoms, particularly alanine, valine or leucine or an oxy-substituted aliphatic amino acid, particularly threonine;

aa^{36} is a neutral aliphatic unsubstituted amino acid of from 3-5 carbon atoms, particularly leucine or a neutral aliphatic oxy-substituted amino acid of from 3-4 carbon atoms, particularly threonine or serine;

aa^{37} is a neutral unsubstituted amino acid, particularly leucine or a thio-substituted amino acid, particularly methionine;

aa^{39} is a neutral aliphatic unsubstituted amino acid, particularly valine or an aromatic unsubstituted amino acid, particularly phenylalanine;

aa^{1} is a neutral aliphatic unsubstituted or oxy-substituted amino acid, particularly alanine, leucine or serine;

aa^{44} is a neutral aliphatic unsubstituted or oxy-substituted amino acid, particularly alanine, isoleucine or threonine;

aa^{46} is a neutral aliphatic unsubstituted amino acid or an oxy-substituted amino acid, particularly alanine, valine or threonine;

aa^{47} is a neutral aliphatic unsubstituted amino acid, particularly glycine or alanine;

aa^{59} is a neutral aliphatic or aromatic unsubstituted amino acid, particularly leucine or phenylalanine;

aa^{76} is a neutral aliphatic unsubstituted or thio-substituted amino acid, particularly alanine, leucine or methionine;

aa^{78} is a neutral aliphatic unsubstituted amino acid, particularly alanine or a neutral aliphatic amino acid having a thio- or an oxy-substitution, particularly methionine or threonine;

aa^{83} is a neutral aliphatic unsubstituted or oxy-substituted amino acid, particularly glycine, serine or threonine;

aa^{92} is a neutral aliphatic amino acid with a oxy-substitution, particularly serine or threonine;

aa^{96} is a neutral aliphatic thio-substituted amino acid or a neutral aromatic heterocyclic amino acid, particularly tryptophan;

aa^{97} is a neutral aliphatic unsubstituted or thio-substituted amino acid, particularly valine, leucine, isoleucine or methionine;

aa^{98} is a neutral aliphatic unsubstituted amino acid or an aromatic oxy-substituted amino acid, particularly alanine, leucine or tyrosine;

aa^{99} may be any amino acid;

aa^{100} is an oxy-substituted amino acid, either aliphatic or aromatic, particularly tyrosine or threonine;

aa^{101} is a neutral unsubstituted aliphatic or aromatic amino acid, particularly alanine, leucine or phenylalanine.

Of particular interest as a source of DNA encoding sequences capable of providing for targeting to an oil body protein are oil-body protein genes obtainable from *Arabidopsis* or *Brassica napus* which provide for expression of the protein of interest in seed (See Taylor *et al.* (1990) *Planta, 181*:18-26). The necessary regions and amino-acid sequences to provide targeting ability to the oil body appear to be the highly hydrophobic central region of oil body proteins.

To identify other oil body protein genes having the desired characteristics, where an oil body protein has been or is isolated, the protein may be partially sequenced, so that a probe may be designed for identifying mRNA. Such a probe is particularly valuable if it is designed to target the coding region of the central hydrophobic domain which is highly conserved among diverse species of plants. In consequence, a DNA or RNA probe for this region may be particularly useful for identifying coding sequences of oil body proteins from other plant species. To further enhance the concentration of the mRNA, cDNA may be prepared and the cDNA subtracted with mRNA or cDNA from non-oil body producing cells. The residual cDNA may then be used for

probing the genome for complementary sequences, using an appropriate library prepared from plant cells. Sequences which hybridize the cDNA under stringent conditions may then be isolated.

In some instances, as described above, using an oil body protein gene probe (conserved region), a probe may be employed directly for screening a cDNA genomic library and identifying sequences which hybridize to the probe. The isolation may also be performed by a standard immunological screening technique of a seed-specific cDNA expression library. Antibodies may be obtained readily for oil-body proteins using the purification procedure and antibody preparation protocol described by Taylor *et al.* (*Planta*, (1990) *181*, 18-26). cDNA expression library screening using antibodies is performed essentially using the techniques of Huynh et al. (1985, in *DNA Cloning, Vol. 1, a Practical Approach*, ed. D.M. Glover, IRL Press, pp. 49-78). Confirmation of sequence is facilitated by the high conservation found in the central hydrophobic region (see Fig. 1). DNA sequencing by the method of Sanger *et al.* (*Proc. Natl. Acad. Sci. USA*, (1977) *74*:5463-5467) or Maxam and Gilbert (1980, *Meth. Enzymol.*, (1980) *65*:497-560) may be performed on all putative clones and homology searches performed. Homology of sequences encoding the central hydrophobic domain is normally ≥ 70%, both at the amino-acid and nucleotide level between diverse species. If an antibody is available, confirmation of sequence identity may also be performed by hybrid-select and translation experiments from seed mRNA preparations as described by Sambrook *et al.* (*Molecular Cloning*, (1990) 2nd Ed., Cold Spring Harbor Press, pp. 8-49 to 8-51).

cDNA clones made from seed can be screened using cDNA probes made from the conserved coding regions of any available oil body protein gene (e.g., Bowman-Vance and Huang (*J. Biol. Chem.*, (1987) *262*:11275-11279). Clones are selected which have more intense hybridization with seed DNAs as compared to seedling cDNAs. The screening is repeated to identify a particular cDNA associated with oil

bodies of developing seeds using direct antibody screening or hybrid-select and translation. The mRNA complementary to the specific cDNA is absent in other tissues which are tested. The cDNA is then used for screening a genomic library and a fragment selected which hybridizes to the subject cDNA.

To obtain expression of the chimeric gene in seed a transcriptional initiation regulatory region and translational initiation regulatory region of untranslated 5' sequences, "ribosome binding sites", responsible for binding mRNA to ribosomes and translational initiation obtainable from any gene preferentially expressed in seed may be used. Examples of such genes include seed storage proteins such as from napin (Josefsson *et al.*, *J. Biol. Chem.*, (1987) *262*:12196-12201; Scofield S.R. and Crouch M.L. *J. Biol. Chem.* (1987) *262*:12202-12208). Preferably, the region is obtainable from an oil body protein (oil-body proteins from *Arabidopsis*, carrot (Hatzopoulos *et al.*, *supra*)or maize (Huang *et al.* 1987 and 1990 supra). The region generally comprises at least 100 bp 5' to the translational start of the structural gene coding sequence, up to 2.5 kb 5' to the same translational start. It is preferred that all of the transcriptional and translational functional elements of the initiation control region are derived from or obtainable from the same gene. By "obtainable" is intended a DNA sequence sufficiently similar to that of a native sequence to provide for the desired specificity of transcription of the DNA sequence encoding the chimeric protein. It includes natural and synthetic sequences and may be a combination of synthetic and natural sequences.

The transcription level should be sufficient to provide an amount of RNA capable of resulting in a modified seed. By "modified seed" is meant seed having a detectably different phenotype from a seed of a non-transformed plant of the same species, for example one not having the expression cassette in question in its genome. Various changes in phenotype are of interest. These changes include

over-expression of oil-body protein or OBP-accumulation on the oil body or in the cytoplasm of the resultant chimeric protein.

The polypeptide of interest may be any protein and includes for example, an enzyme, an anticoagulant, a neuropeptide, a hormone, or adhesive precursor. Examples of proteins include interleukin-1-β, the anticoagulant Hirudin, the enzyme β-glucuronidase or a single-chain antibody comprising a translational fusion of the V_H or V_L chains of an immunoglobin. The DNA sequence encoding the polypeptide of interest may be synthetic, naturally derived, or combinations thereof. Depending upon the nature or source of the DNA encoding the polypeptide of interest, it may be desirable to synthesize the DNA sequence with plant preferred codons. The plant preferred codons may be determined from the codons of highest frequency in the proteins expressed in the largest amount in the particular plant species of interest as a host plant.

The termination region which is employed will be primarily one of convenience, since in many cases termination regions appear to be relatively interchangeable. The termination region may be native with the transcrip-tional initiation region, may be native with the DNA sequence encoding the polypeptide of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions.

Ligation of the DNA sequence encoding the targeting sequence to the gene encoding the peptide of interest may take place in various ways including terminal fusions, internal fusions, and polymeric fusions. In all cases, the fusions are made so as not to interrupt the reading frame of the oil-body protein and so as to avoid any translational stop signals in or near the junctions. The different types of terminal and internal fusions are shown in Fig. 2 along with a representation of their configurations *in vivo*.

In all the cases described, the ligation of the gene encoding the peptide preferably would include a linker encoding a protease target motif. This would permit the release of the peptide once extracted as a fusion protein. Potential cleavage sites which could be employed are recognition motifs for thrombin (leu-val-pro-arg-gly) (Fujikawa *et al.*, *Biochemistry* (1972) *11*:4892-4899), of factor Xa (phe-glu-gly-arg-aa.) (Nagai *et al.*, *Proc. Nat'l Acad. Sci. USA*, (1985) *82*:7252-7255) or collagenase (pro-leu-gly-pro) (Scholtissek and Grosse *Gene* (1988) 62:55-64).

By appropriate manipulations, such as restriction, chewing back or filling in overhangs to provide blunt ends, ligation of linkers, or the like, complementary ends of the fragments can be provided for joining and ligation. In carrying out the various steps, cloning is employed, so as to amplify the amount of DNA and to allow for analyzing the DNA to ensure that the operations have occurred in proper manner. A wide variety of cloning vectors are available, where the cloning vector includes a replication system functional in *E. coli* and a marker which allows for selection of the transformed cells. Illustrative vectors include pBR332, pUC series, M13mp series, pACYC184, etc. Thus, the sequence may be inserted into the vector at an appropriate restriction site(s), the resulting plasmid used to transform the *E. coli* host, the *E. coli* grown in an appropriate nutrient medium and the cells harvested and lysed and the plasmid recovered. Analysis may involve sequence analysis, restriction analysis, electrophoresis the like. After each manipulation the DNA sequence to be used in the final construct may be restricted and joined to the next sequence, where each of the partial constructs may be cloned in the same or different plasmids.

A variety of techniques are available for the introduction of DNA into plant cell host. For example, the chimeric DNA constructs may be introduced into host cells obtained from dicotyledenous plants, such as tobacco, and oleaginous species, such as Brassica napus using standard *Agrobacterium* vectors by a transformation protocol such as

that described by Moloney *et al. Plant Cell Rep.*, (1989) *8*:238-242 or Hinchee *et al. Bio/Technol.*, (1988) *6*:915-922; or other techniques known to those skilled in the art. For example, the use of T-DNA for transformation of plant cells has received extensive study and is amply described in EPA Serial No. 120,516; Hoekema, In: The Binary Plant Vector System Offset-drukkerij Kanters B.V., Alblasserdam, 1985, Chapter V, Knauf, *et al.*, *Genetic Analysis of Host Range Expression by Agrobacterium*, In: Molecular Genetics of the Bacteria. Plant interaction, Puhler, A. ed., Springer-Verlag, NY, 1983, p. 245, and An *et al.*, *EMBO J.* (1985), *4*:277-284. Conveniently, explants may be cultivated with *A. tumefaciens* or *A. rhizogenes* to allow for transfer of the transcription construct to the plant cells. Following transformation using Agrobacteria the plant cells are dispersed in an appropriate selective medium for selection, grown to callus, shoots grown and plantlets regenerated from the callus by growing in rooting medium. The *Agrobacterium* host will contain a plasmid having the *vir* genes necessary for transfer of the T-DNA to the plant cells and may or may not have T-DNA. For injection and electroporation, (see below) disarmed Ti-plasmids (lacking the tumor genes, particularly the T-DNA region) may be introduced into the plant cell.

The use of non-*Agrobacterium* techniques permits the use of the constructs described herein to obtain transformation and expression in a wide variety of monocotyledonous and dicotyledonous plants. These techniques are especially useful for species that are intractable in an *Agrobacterium* transformation system. Other techniques for gene transfer include biolistics (Sanford, *Trends in Biotech.* (1988) *6*:299-302), electroporation (Fromm *et al.* (1985) *Proc. Nat'l. Acad. Sci. USA*, *82*:5824-5828; Riggs and Bates (1986), *Proc. Nat'l. Acad. Sci. (USA)* 83 5602-5606 or PEG-mediated DNA uptake (Potrykus *et al.* (1985) *Mol. Gen. Genet.*, *199*:169-177).

As a host cell, cells from any of a number of seed bearing plants may be employed in which the cells are derived from plant parts such as stem, leaf, root, or seed or reproductive structures according to the species. The cells may be isolated cells or plant parts, for example, leaf discs. In a specific application, such as to *Brassica napus*, the host cells generally will be derived from cotyledonary petioles as described by Moloney *et al. Plant Cell Rep.*, (1989) *8*:238-242). Other examples using commercial oil seeds include cotyledon transformation in soybean explants (Hinchee *et al. Biotechnology*, (1988) *6*:915-922) and stem transformation of cotton (Umbeck *et al. Biotechnology*, (1981) *5*:263-266).

Following transformation, the cells, for example as leaf discs, are grown in selective medium. Once shoots begin to emerge, they are excised and placed onto rooting medium. After sufficient roots have formed, the plants are transferred to soil. Putative transformed plants are then tested for presence of a marker. Southern blotting is performed on genomic DNA using an appropriate probe, for example an *A. thaliana* oleosin gene, to show that integration of the desired sequences into the host cell genome has occurred.

The expression cassette will normally be joined to a marker for selection in plant cells. Conveniently, the marker may be resistance to a herbicide, particularly an antibiotic, such as kanamycin, G418, bleomycin, hygromycin, chloramphenicol, or the like. The particular marker employed will be one which will allow for selection of transformed cells as compared to cells lacking the DNA which has been introduced.

The fusion peptide in the expression cassette constructed as described above, expresses at least preferentially in developing seeds. Accordingly, transformed plants grown in accordance with conventional ways, are allowed to set seed. See, for example, McCormick *et al.* Plant Cell Reports (1986) 5:81-84. Northern blotting can be carried out using an appropriate gene probe with RNA

isolated from tissue in which transcription is expected to occur such as a seed embryo. The size of the transcripts can then be compared with the predicted size for the fusion protein transcript.

Oil-body proteins are then isolated from the seed and analyses performed to determine that the fusion peptide has been expressed. Analyses can be for example by PAGE. The fusion peptide can be detected using an antibody to the oleosin portion of the fusion peptide. The size of the fusion peptide obtained can then be compared with predicted size of the fusion protein.

Two or more generations of transgene plants may be grown and either pollinated with the same transformed strain or different strains, identifying the resulting hybrid having the desired phenotypic characteristic, to ensure that the subject phenotypic characteristic is stably maintained and inherited and then seeds harvested for isolation of the peptide of interest or for use to provide seeds with the new phenotypic property.

The desired protein can be extracted from seed that is homo- or heterozygous for the introduced trait by a variety of techniques, including use of an aqueous, buffered extraction medium and a means of grinding, breaking, pulverizing or otherwise disrupting the cells of the seeds. The extracted seeds can then be separated (for example, by centrifugation or sedimentation of the brei) into three fractions: a sediment or insoluble pellet, an aqueous supernatant, and a buoyant "scum" comprising seed storage lipid and oil bodies. These oil bodies contain both native oil-body proteins and chimeric oil body proteins, the latter containing the foreign peptide. The oil-bodies are separated from the water-soluble proteins and re-suspended in aqueous buffer.

If a linker comprising a protease recognition motif has been included in the expression cassette, to the resuspension buffer is added a protease specific for the recognition motif produced by translation of the linker sequence. This releases the required peptide into the

aqueous phase. A second centrifugation step will now re-float the processed oil-bodies with their attached proteins and leave an aqueous solution of the required peptide. The desired peptide may be precipitated, chemically modified or lyophilized according to its properties and desired applications

In certain applications it may not be necessary to remove the chimeric protein from the oil-body protein. Such an application would include cases where the fusion peptide includes an enzyme which is tolerant to N or C-terminal fusions and retains its activity; such enzymes could be used without further cleavage and purification. The chimeric enzyme-OBP would be contacted with substrate as a fusion protein. It is also possible, if desired, to purify the enzyme - OBP fusion protein using an immunoaffinity column comprising an immobilized high titre antibody against the OBP (see, for example, Taylor *et al.*, (1990) supra).

Other uses for the subject invention are as follows. OBP's comprise a high percentage of total seed protein, thus it is possible to enrich the seed for certain desirable properties such as high-lysine, high methionine, and the like, simply by making the fusion protein rich in the amino-acid(s) of interest could find utility of particular interest is the modification of grains and cereals which are used as either directly or indirectly as food sources for livestock, including cattle, poultry, and humans. It may be possible to include, as the fusion peptide, an enzyme which may assist in subsequent processing of the oil or meal in conventional oilseed crushing and extraction, for example inclusion of a thermostable lipid-modifying enzyme which would remain active at the elevated crushing temperatures used to process seed and thus add value to the extracted triglyceride or protein product. Other uses of the fusion protein to include use to improve the agronomic health of the crop. For example, an insecticidal protein or a portion of an immunoglobulin specific for an agronomic pest such as a fungal cell wall or

membrane, could be coupled to the oil body protein thus reducing attack of the seed by a particular plant pest.

The following examples are offered by way of illustration and not by limitation.

EXPERIMENTAL

Example 1
Expression of terminal fusions of foreign peptides with oil-body proteins

A. C-terminal fusions

A genomic clone of an oil-body protein gene containing at least 100 bp 5' to the translational start is cloned into a plasmid vehicle capable of replication in a suitable bacterial host (e.g., pUC or pBR322 in *E. coli*). A restriction site is located in the region encoding the hydrophilic C-terminal portion of gene. In a 19 kDa OBP, this region stretches typically from codons 125 to the end of the clone. The ideal restriction site is unique, but this is not absolutely essential. If no convenient restriction site is located in this region, one may be introduced by the site-directed mutagenesis procedure of Kunkel *Proc. Nat'l. Acad. Sci. USA*, (1985) *82*:488-492. The only major restriction on the introduction of this site is that it must be placed 5' to the translational stop signal of the OBP clone.

With this mutated clone in place, a synthetic oligonucleotide adapter may be produced which contains coding sequence for a protease recognition site such as Pro-Leu-Gly-Pro or a multimer thereof. This is the recognition site for the protease collagenase. The adaptor would be synthesized in such a way as to provide: a 4-base overhang at the 5' end compatible with the restriction site at the 3' end of the OBP clone, a 4-base overhang at the 3' end of the adaptor to facilitate ligation to the foreign peptide coding sequence and additional bases, if needed, to ensure no frame shifts in the transition between the OBP

coding sequence, the protease recognition site and the foreign peptide coding sequence. A typical arrangement for such a fusion is shown in Fig. 3. The example shown here uses an existing *Xho*1 site near the stop codon of a carrot OBP (Hatzopoulos *et al. Plant Cell*, (1990) *2*:457-467). This is digested and may be ligated with an adapter constructed from the two oligonucleotides described. This adapter will form a perfect *Xho*1 overhang at an end and will not disrupt the translational frame. The other end forms an *Nco*1 overhang which is arbitrarily chosen (any six-base cutter will suffice), but which encloses an ATG from the desired foreign peptide.

The final ligation product will contain an almost complete OBP gene, coding sequence for collagenase recognition motif and the desired peptide coding region all in a single reading frame. This tripartite fragment is cloned into an *Agrobacterium* binary plasmid (Bevan *Nucl. Acid Res.*, (1984) *12*:8711-8721) such as is widely used to transfer foreign DNA into plants (Fraley *et al. Proc. Nat'l Acad. Sci. USA*, (1983) *80*:4803-4807) and this is used to transform oilseed plants such as rapeseed using the method of Moloney *et al. Plant Cell Rep.*, (1989) *8*:238-242) or similar procedure. Transgenic plants may be recovered from this transformation experiment and these are grown to flowering. The plants then set seed by self-fertilization.

The seeds are allowed to reach maturity (60-80 days) and then are harvested and ground in aqueous extraction buffer (Taylor *et al. Planta*, (1990) *181*:18-26). The slurry is centrifuged at 5000 xg for 20 min. and will give a surface scum. This scum is again recovered and suspended by vigorous shaking in a collagenase assay buffer (Scholtissek and Grosse, *Gene* (1988) *62*:55-64). Five units of collagenase are added and the suspension is incubated with shaking for 4 h. After this time, the suspension is once again centrifuged at 5000 xg for 20 min. The surface scum is removed and the protein content of the aqueous phase is analyzed by SDS-Poly Acrylamide Gel Electrophoresis. If

a band of approximately the size of the required peptide is found, the protein may be precipitated using ammonium sulfate, concentrated using ultrafiltration or lyophilized.

B. N-terminal fusions

The hydrophilic N-terminal end of oil-body proteins permits the fusion of peptides to the N-terminal while still assuring that the foreign peptide would be retained on the outer surface of the oil body. The configuration of such fusions is shown in Fig. 2IB.

This configuration can be constructed from similar starting materials as used for C-terminal fusions, but requires the identification of a convenient restriction site close to the translational start of the oil-body protein gene. A convenient site may be created in many oil-body protein genes without any alteration in coding sequence by the introduction of a single base change just 5' to the first 'ATG'. In oil body proteins thus far studied, the second amino acid is alanine whose codon begins with a "G". The context of the sequences is shown below:

A-C transition here yields *Nco*I site

↓

3' . . . TC TCA ACA ATG GCA . . . Carrot OBP

3' . . . CG GCA GCA ATG GCG . . . Maize 18KDa OBP

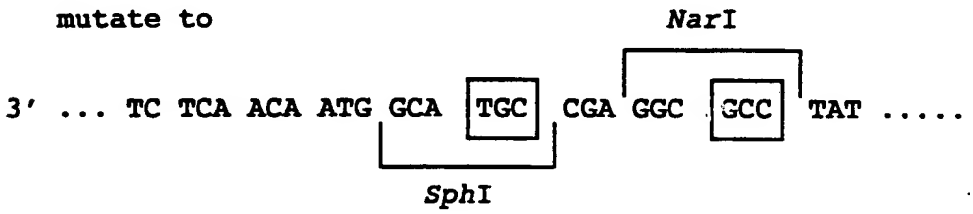
A single base change at the adenine prior to the 'ATG' would yield in both cases . . . CCATGG . . . which is an *Nco*1 site. Thus, modification of this base using the site-directed mutagenesis protocol of Kunkel (*Proc. Nat'l. Acad. Sci. USA*, (1985) *82*:488-492) will prepare this clone for use assuming no other *Nco*1 sites in the sequence.

The coding sequence for the foreign peptide may require preparation which will allow its ligation directly into the *Nco*1 site. This may typically require a single or two-base modification by site-directed mutagenesis (Kunkel, 1985, *supra*) to generate an *Nco*1 site around the

translational start of the foreign peptide. This peptide is then excised from its cloning vehicle using *Nco*1 and a second enzyme which cuts close to the translational stop of the target. Again, using the methods described above, a second convenient site can be introduced by site-directed mutagenesis. It has been suggested by Qu and Huang (1990, *supra*) that the N-terminal methionine might be removed during processing of the protein *in vivo* and that the alanine immediately downstream of this might be acylated. To account for this possibility, it may be necessary to retain the Met-Ala sequence at the N-terminal end of the protein. This is easily accomplished using a variety of strategies which introduce a convenient restriction site into the coding sequence in or after the Ala codon. For example, by site-directed mutagenesis, the sequences could be modified as follows:

3' ... TC TCA ACA ATG GCA GAA CGA GGC ACT TAT .....

mutate to



This change of a single codon would introduce a *Sph*1 site into the coding sequence. A second change, which could be introduced during the same round of mutagenesis would convert two bases in codon 6 to yield GGC GCC, an *Nar*1 site. This mutated gene could then be opened with *Sph*1 and *Nar*1 to give a directional cloning cut which would eliminate three codons. Into this site could be introduced an adaptor containing a 3' overhang with the sequence CATG ... (compatible with *Sph*1) and a GC 5' overhang at the opposite end. The precise sequence of this adapter is shown below:

*Sph*1 *Nar*1

n times

CTACG CCG CTC GGT CCG GG
GGC GAG CCA GGC CCGC

This adapter would recreate both the *Sph*1 and *Nar*1 restriction sites which would be used for diagnostic purposes. The *Sph*1 site could now be used to open the plasmid and clone in-frame a DNA fragment enclosing the sequence for a useful peptide. Orientation of cloning could then be analyzed by cutting at any asymmetrically placed site and *Nar*1 of the plasmid.

The resultant constructs from these N-terminal fusions would be typical of the examples IB of Figure 2. They would contain an OBP promoter sequence, an in-frame fusion in the first few codons of the OBP gene of a high value peptide coding sequence with its own ATG as start signal if necessary and the remainder of the OBP gene and terminator.

This modified gene is introduced into a binary *Agrobacterium* plasmid (Bevan, (1984), *supra*) and mobilized into *Agrobacterium*. Transformations are performed as described above. Recovery of the high value peptide from seeds is performed as described for 'C-terminal fusions.'

C. Internal translational fusions

A third type of fusion involves the placing of a high value peptide coding sequence internally to the coding sequence of the OBP. This type of fusion requires the same strategy as in N-terminal fusions, but may only be functional with modifications in regions of low conservation, as it is believed that regions of high conservation in these OBPs are essential for targeting of the mature protein.

The key difference in this kind of fusion is the necessity for *flanking* collagenase recognition sites for the release of the protein. This means that in place of the standard collagenase linker/adapter systems thus far

described, it is necessary to have a linker with the following form:

	n times		n times	
Cohesive end 1	CCG CTC GGT CCG GGC GAG CCA GGC	Restriction site	CCG CTC GGT GGC GAG CCA GGC	end 2

Cohesive ends 1 and 2 would be used to clone the adapter into an OBP clone in a directional manner. The nested restriction site is then used to introduce the high value peptide coding sequence flanked by appropriate restriction sites or linkers. Orientation is checked by the use of an asymmetrically placed restriction site in the high-value peptide coding sequence and one of the two restriction sites flanking the coding sequence of the collagenase recognition motif.

Mobilization of these constructs to *Agrobacterium* plasmids and then to plants is identical to the previous described procedure. Recovery of the high-value protein from the seeds of transgenic plants is somewhat different in that after the oil-bodies have been isolated and washed, it may be necessary to delipidate the oil-bodies in order to access the collagenase recognition sites which could be hidden inside the oil-body in the lipid phase. This step may reduce certain advantages of using oil-body proteins as carriers, but may on the other hand be very convenient for protein sequences which are labile in aqueous media or in plant cytoplasms.

D. Inter-dimer translational fusions

It is possible to create a construct in which the entire coding sequence of the OBP is repeated. A dimeric protein produced from this construct may still contain all the necessary factors for targeting the OBP to the oil-body. Such a construct would contain a promoter region, an entire or near-complete open reading frame for an OBP but excluding

the translational stop and then an entire open-reading frame of a second OBP, this time equipped with a translational 'stop' and a terminator region.

In the construction of this chimeric gene a pair of dissimilar restriction sites are either found or created at the region of the junction of the two copies. These sites are used to enable the introduction of a linker such as is described above for internal translational fusions. The linker contains not only sets of collagenase recognition motifs, but also an internal restriction site in which to nest a sequence encoding a high value protein. The form of this construct is shown in Fig. 2 III. Mobilization of this construct to *Agrobacterium* and then to plants is exactly as above. Recovery of the high value protein from seeds of the transformed plants would be performed using the same procedure as described for C-terminal fusions above.

Example 2
Strategy for the cloning and expression of Interleukin-1-β (IL-1-β) as a fusion with oleosins in plants

A. Cloning and sequencing of an *Arabidopsis thaliana* oleosin gene

A *Brassica napus* oleosin gene (Murphy *et al*, (1991) *Biochim Biophys Acta 1088*: 86-94) was used to screen a genomic library of *A. thaliana* (cv. Columbia) in EMBL3A (Stratagene). The screening resulted in the isolation of a EMBLA3A clone (λ2.1) containing a 15 kb genomic fragment from *A. thaliana*. The oleosin was mapped within a 6.6 kb *Kpn*I insert, within this 15 kb fragment (Fig. 5). A 1.8 kb *Nco*I/*Kpn*I fragment containing the oleosin gene was end filled and subcloned in the *Sma*I site of RFM13mp19. The 1.8 kb insert was digested with convenient restriction enzymes and subcloned in M13mp19 for sequencing. The 1800 bp sequence of the *A. thaliana* oleosin gene is presented in Fig. 1a. All the cloning procedures were carried out according to Sambrook *et al.*, (1989) (Molecular Cloning: A

laboratory manual 2nd ed. Cold Spring Haber Laboratory Press.)

B. Design of an oligonucleotide encoding IL-1-β

IL-1-β consists of 9 amino acids (aa); val-gln-gly-glu-glu-ser-asn-asp-lys (Antoni *et al.*, (1986) *J. Immunol.* *137*:3201-3204). The protease factor Xa can cleave a protein sequence which contains a aa sequence ile-glu-gly-arg. Cleavage takes place after the aa arg. Based on these sequences an oligonucleotide was designed (GVR11, fig. 5), which contains in addition to the IL-1-β coding sequence, the coding sequence for the factor Xa cleavage site, and 18 nucleotides of the 3' coding region of the *A. thaliana* oleosin (base position 742-759). The IL-1-β coding sequence was designed using optimal codon usage for the *B. napus* and *A. thaliana* oleosin (Table 1).

Table 1.

Codon usage of A. thaliana[1] and B. napus oleosin[2]

TTT	phe	F	5	TCT	ser	S	8	TAT	tyr	Y	6	TCT	cys	C	-
TTC	phe	F	3	TCC	ser	S	10	TAC	tyr	Y	11	TGC	cys	C	-
TTA	leu	L	-	TCC	ser	S	2	TAA	OCH	z	-	TGA	OPA	z	-
TTG	leu	L	4	TGA	ser	S	-	TAG	AMB	z	-	TGG	trp	w	2
CTT	leu	L	10	CCT	pro	p	3	CAT	his	H	4	CGT	arg	R	6
CTC	leu	L	11	CCC	pro	p	1	CAC	his	H	6	CGC	arg	R	-
CTA	leu	L	-	CCA	pro	p	4	CAA	gln	Q	4	CGA	arg	R	1
CTG	leu	L	6	CCG	pro	p	2	CAG	gln	Q	17	CGG	arg	R	-
ATT	ile	I	7	ACT	thr	T	11	AAT	asn	N	-	AGT	ser	S	5
ATC	ile	I	13	ACC	thr	T	14	AAC	asn	N	1	AGC	ser	S	3
ATA	ile	I	3	ACA	thr	Y	7	AAA	lys	K	5	AGA	arg	R	5
ATG	met	M	11	ACG	thr	T	5	AAG	lys	K	10	AGG	arg	R	3
GTT	val	V	11	GCT	ala	A	17	GAT	asp	D	8	GGT	gly	G	0
GTC	val	V	9	GCC	ala	A	2	GAC	asp	D	19	GGC	gly	G	9
GTA	val	V	-	GCA	ala	A	10	GAA	glu	E	2	GGA	gly	G	14
GTG	val	V	-	GCG	ala	A	2	GAG	gly	E	2	GGG	gly	G	5

1 See Figure 1A.
2 Lee and Huang (1991) Plant Physical 96: 1395-1397.

C. Creation of an *A. thaliana* oleosin-IL-1-β fusion

Based on the sequence:

5' CACACCAGGAACTCTCTGGTAAGC 3'

(base position: -838 to -814), oligonucleotide GVR10

5' CACTGCAGGAACTCTCTGGTAAGC 3'

was designed. GVR10 contains a *Pst*I restriction site (underlined) to facilitate cloning. The polymerase chain reaction (PCR) was used amplify the region between GVR10 and GVR11. The reaction mixture contained: 16 μl dNTPs (1.25 mM), 10 μl 10X PCR buffer (100 mM Tris-HCL pH 8.3, 500 mM KCL, 15 mM $MgCl_2$, 0.1% (w/v) gelatin), 5 μl GVR11 (20 μM) 1 μl Taq DNA polymerase (1 u/μl) and 64 μl H_2O. The reaction was carried out for 30 cycles. Each cycle consisted of 1 minute denaturing at 92°C, 1 minute annealing at 45°C and 3 minutes extension at 72°C. The PCR reaction yielded a single fragment of 1652 nucleotides.

D. Cloning of the *A. thaliana* oleosin-IL-1-β (OBPIL) fusion

A 5' SalI-nopaline synthase (nos) terminator-*Eco*RI 3' sequence was isolated from pBI121 (Clontech laboratories) and cloned into the *Sal/Eco*RI sites of pUC19. The plasmid was called pTerm. The 1652 bp fragment (described in C.) was isolated and digested with the restriction enzymes *Pst*I and *Sal*I. This fragment was cloned in pTerm. The resulted plasmid was called pUCOBPILT (fig. 5). This plasmid was digested with *Eco*RI and *Pst*I and resulted in the digested pUC19 vector and the *Eco*RI-*A. thaliana* oleosin-IL-1-β -nos -*Pst*I fusion *Pst*I (OBPILT). The complete sequence of OBPILT is shown in fig. 7. OBPILT was subcloned in the *Eco*RI/*Pst*I sites of pBluescript+. This plasmid (pBIOBPILT) was digested with *Pst*I and *Hind*III and the *Pst*I-OBPILT-*Hind*III fragment was subcloned in a binary *Agrobacterium* plasmid (Bin 19) (Bevan, M., (1984) *Nucl. Acid. Res.* 12: 8711-8721) containing a selection marker (neomycin phosphotransferase and *Pst*I-*Hind*III unique sites. The resulting plasmid was called pCGOBPILT. A schematic

representation of the cloning procedure is shown in Fig. 5. For descriptions of various binary plasmids, see, pGA642 or 645; An *et al.* (1985) EMBO. J. 4 277-288 or pCGN1558 or 1559; MacBride and Summerfeldt (1990) *Plant. Molec. Biol.* 14 269-276.

F. Transformation of pCGOBPILT into *Agrobacterium* strain EHA101

A single EHA101 colony (Hood *et al.*, (1986) *J. Bact. 168*:1291-1301) was used to inoculate 5 ml of LB+100 µg/ml kanamycin. This culture was grown for 48 hours at 28°C. This 5 ml culture was used to inoculate 500 ml of LB+100 µg/ml kanamycin. This culture was grown at 28°C until the culture reached a density of OD600=0.5 (approx. 4 hours). The cells were spun down (10 min, 5000 x g) and resuspended in 500 ml of sterile H_2O (repeated 2x). The cells were spun again and resuspended in 3 ml sterile H_2O, containing 10% glycerol. 40 µl of the cells were aliquoted in Eppendorf tubes and either directly used for electroporation, or stored at -80°C for future use. Electroporation was carried out according to Bower *et al.*, *Nucl. Acid. Res.* (1988) 16 6127-6145. The pulse generator was set to the 25 µF capacitor, 2.5 kV and 200 ohm in parallel with the sample chamber.

G. Transformation of *Nicotiana tabacum* (tobacco) with pCGOBPILT

The EHA 101 containing pCGOBPILT was used to transform tobacco leaf discs. Eight to ten centimeter long tobacco leaves were taken from greenhouse grown plants, sterilized in 70% ethanol for 20 sec. and then in 10% bleach (such as Javex) for 8 min. The leaves were then rinsed 6 times with sterile water. The leaf edges as well as the midrib were excised from the leaves and the remaining lamina was sectioned into 5x7 mm squares or discs of 5 mm diameter. About 30 leaf discs were collected and placed into a small petri dish. The *Agrobacterium* solution was then poured over the tobacco discs and incubation occurred for 9 minutes.

The leaf pieces were then blotted on sterile Whatman filter paper and placed, abaxial side down, onto media I (MS, 3% sucrose and 2 mg/l 2,4-D). Co-cultivation proceeded for the following 48 hours. At this point the leaf discs were transferred to selection media (MD, 3% sucrose, 2.5 mg/l Ba, 0.1 mg/l NAA, 500 mg/l carbenicillin, and 100 mg/l kanamycin) where they remained for the next 3-4 weeks. Once shoots began to emerge they were excised and placed onto rooting media (MS, 3% sucrose, 0.1 mg/l NAA, 500 mg/l carbenicillin, and 50 mg/l kanamycin). After sufficient roots had been formed, the tobacco plants were transferred to soil.

H. Transformation of *B. napus* with pCGOBPILT

The transformation of *B. napus* was carried out according to Moloney *et al.*, (1989) *Plan Cell Rep, 8*:238-242, which disclosure is incorporated herein by reference.

Transformation Procedure

Single colonies of *Agrobacterium tumefaciens* strain EHA 101 containing the binary plasmid were grown overnight at 28°C in AB medium. A 50µl sample of this suspension was grown overnight at 28°C in 5 ml of MG/L broth supplemented with appropriate antibiotics. This bacterial suspension was pelleted by centrifugation for 15 min. at 10,000 x g then resuspended in 10 ml of MS medium containing 3% sucrose and at pH 5.8. A thin film of this suspension was used to cover the base of a 5 cm petri dish. Individual excised cotyledons were taken from the plates described above and the cut surface of their petioles was immersed into this bacterial suspension for a few seconds. They were immediately returned to the same MS plates from which they had been taken. The cotyledons were co-cultivated with the *Agrobacterium* for 72 h. No feeder layers were employed.

After co-cultivation, the cotyledons were transferred to regeneration medium comprising MS medium supplemented with 20µM benzyladenine, 3% sucrose, 0.7% phytagar, pH 5.8 and 500 mg/l carbenicillin (Pyopen,

Ayerst) and 15 mg/l kanamycin sulphate (Boehringer-Mannheim). Again the petioles were carfully embedded in the agar to a depth of 2 mm. Plating density was maintained at 10 explants per plate. Higher densities reduce regeneration frequency.

Selection and Plant Regeneration

The explants were maintained on regeneration medium under light and temperature conditions specified above for 2-3 weeks. During this time many shoots appeared on over half the explants with relatively little callus formation. Some of these shoots undergo bleaching by the fourth week of culture. The remaining green shoots were subcultured onto shoot elongation medium which was the same as regeneration medium minus the benzyladenine. One or two weeks on this medium permitted the establishment of apical dominance from the shoot clusters formed. The shoots so derived were transferred to "rooting" medium containing MS medium, 3% sucrose, 2mg/l indole butyric acid, 0.7% phytagar and 500 mg/l carbenicillin. No kanamycin was used at this stage as it was found that more rapid root establishment occurred without the selection agent while very few "escapes" actually succeeded in rooting after the two rounds of selection on regeneration and shoot elongation medium.

I. **Stable integration of OBPILT in the tobacco and *B. napus* genomes**

Putative transformed plants were tested for neomycin phosphotransferase activity. Genomic DNA from plants showing this activity was isolated. Southern blotting was performed in order to demonstrate that the sequences between the T-DNA borders (OBPILT and neomycin phosphotransferase gene) were stably integrated in to the genomes of *B. napus* and tobacco. The tobacco Southern was probed with the *A. thaliana* oleosin gene, and the neomycin

phosphotransferase gene. The *B. napus* Southern was probed with the neomycin phosphotransferase gene.

J. Expression of the oleosin-IL-1-β fusion in tobacco plants

RNA was isolated from developing embryos obtained from transformed and untransformed plants. Northern blotting was carried out using the *A. thaliana* oleosin as a gene probe. In all the tested transformed plants a 850 nt transcript could be detected. The size of these transcripts correspond to the expected size of the oleosin-IL-1-β mRNA. These transcripts could not be detected in the untransformed plants.

K. Accumulation of the oleosin-IL-1-β protein

Oil-body proteins were isolated from transformed tobacco seeds (Holbrook *et al.*, (1991) *Plant Physical* 97:1051-1058. PAGE was performed and the protein were transferred from the gel to PVDF membranes. An antibody, which was raised against a 22 kDa oleosin of *B. napus*, was used to detect the oleosin-IL-1-β fusion in the tobacco seeds. This antibody recognizes all the major oleosins in *B. napus* and *A. thaliana*. In addition, this antibody recognizes the tobacco oleosins. Tobacco oleosins have different sizes from the *A. thaliana* and *B. napus* oleosins. In the transformed tobacco seeds the anti-22 kDa antibody recognized a 20 kDa-protein, which was not present in the untransformed tobacco seed. The predicted size of the oleosin-IL-1-β fusion is 20.1 kDa. A summary of the results is shown in Table 2.

By expressing a peptide of interest conjugated to an oil body protein, or a sufficient portion thereof to provide for getting to the oil bodies, the peptide of interest can be easily purified so as to be substantially free of other cellular components. The fusion protein can be cleaved following purification or may be used without cleavage. The subject methods and compositions provide a fast, simple method for purifying a polypeptide of i

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

Table 2.

Summary of expression results[1]

	Stable integration Southern Blotting	Expression of Oleosin -IL-1b Northern Blotting	Detection of 20kDa protein Western Blotting
B. napus:			
not transformed	-	NT	NT
transformant #12	+	NT	NT
Tobacco:			
not transformed	-	+	+
transformant A	+	+	+
transformant B	+	+	+
transformant G	+	+	+
transformant L	+	+	+

[1] Stable integration, determined as described in I, supra. Expression of the oleosin-Il-1-b fusion as described in J, supra. Detection of a putative oleosin -IL-1-b fusion protein was as described in K, supra. NT=Not Tested

WHAT IS CLAIMED IS:

1. A polypeptide, characterized as capable of targeting an oil body, and having the formula

$$pp^{1}- aa^{25}- aa^{26}- \begin{matrix} V \\ A \end{matrix} - \begin{matrix} V- \\ A \end{matrix} T - \begin{matrix} L- \\ A \end{matrix} aa^{21}- \begin{matrix} P- \\ T \end{matrix}$$

$$aa^{34}- G- G- aa^{36} - \begin{matrix} L- \\ M \end{matrix} L- aa^{39}- L- aa^{41}-$$

$$\begin{matrix} G \\ S \end{matrix} - \begin{matrix} I \\ L \end{matrix} - aa^{44}- L- aa^{46}- aa^{47}- \begin{matrix} T \\ S \end{matrix} - \begin{matrix} L \\ V \end{matrix} - \begin{matrix} I \\ V \end{matrix} -$$

$$aa^{51}- L- aa^{53}- V- A- T- P- \begin{matrix} L \\ V \end{matrix} - aa^{59} - \begin{matrix} L \\ V \end{matrix} -$$

$$\begin{matrix} L \\ I \end{matrix} - F- S- P- \begin{matrix} V \\ I \end{matrix} - L- V- P- A- \begin{matrix} A \\ L \end{matrix} - \begin{matrix} L \\ I \end{matrix} - aa^{73} -$$

$$aa^{74} - aa^{75} - aa^{76} - aa^{77} - aa^{78} - G- F- L-$$

$$\begin{matrix} G \\ S \\ T \end{matrix} - S- aa^{87} - G- \begin{matrix} L \\ V \\ I \end{matrix} - aa^{89} - aa^{90} - \begin{matrix} L \\ I \end{matrix} - \begin{matrix} S \\ T \end{matrix} -$$

$$aa^{93} \quad aa^{94} - \begin{matrix} S \\ T \end{matrix} - aa^{96} - aa^{97} - aa^{98} - aa^{99} -$$

$$aa^{100} - aa^{101} - pp^{2}$$

with the proviso that said peptide is other than the full length naturally occurring 16Kd oleosin from carrot or the 18 Kd or 16 Kd oleosin from maize.

2. The polypeptide according to Claim 1, wherein at least one of PP^1 and PP^2 comprises a polypeptide of interest.

3. The polypeptide according to Claim 1, wherein at least one of PP^1 and PP^2 comprises an antigenic amino acid sequence to provide an immunogen.

4. A polypeptide characterized as capable of targeting an oil body selected from the group consisting of:

(a) a peptide comprising at least eight consecutive amino acids included in the following amino acid sequence:

10 20
M-N-G-R-D-R-D-Q-Y-Q-M-S-G-R-G-S-D-Y-S-K-
30 40
S-R-Q-I-A-K-A-A-T-A-V-T-A-G-G-S-L-L-V-L-
50 60
L-S-L-T-L-V-G-T-V-I-A-L-T-V-A-T-P-L-L-V-
70 80
I-S-S-T-I-L-V-P-A-L-I-T-V-A-L-L-I-T-G-S-
90 100
L-S-S-G-G-F-G-I-A-A-I-T-V-F-S-W-I-Y-K*Y-
110 120
L-L-I-E-H-P-Q-G-S-D-K-L-D-S-A-R-M-K-L-G-
130 140
S-K-A-Q-D-L-K-D-R-A-Q-Y-Y-G-Q-Q-H-T-G-W-
150
E-H-D-R-D-R-T-R-G-G-Q-H-T-T; and

(b) a peptide that is encoded by a DNA sequence identified by means of an oligonucleotide probe designed based upon said amino acid sequence in (a) or a fragment thereof, with the proviso that said peptide is other than the full length naturally occurring 16Kd oleosin from carrot or the 18 Kd or 16 Kd oleosin from maize.

5. The polypeptide according to Claim 4, wherein said peptide in (a) comprises at least twelve consecutive amino acids included in the following amino acid sequence:

40

30 40
A-K-A-A-T-A-V-T-A-G-G-S-L-L-V-L-
50 60
L-S-L-T-L-V-G-T-V-I-A-L-T-V-A-T-P-L-L-V-
70 80
I-S-S-T-I-L-V-P-A-L-I-T-V-A-L-L-I-T-G-S-
90 100
L-S-S-G-G-F-G-I-A-A-I-T-V-F-S-W-I-Y-K*Y-
101
L.

6. A DNA construct comprising a DNA sequence encoding (a) an oleosin or a portion thereof sufficient to provide for targeting to an oil body and (b) a polypeptide of interest.

7. The DNA construct according to Claim 6, further comprising:

vector DNA containing at least one regulatory sequence operatively associated with said DNA sequence which is capable of directing replication of said DNA in a host cell.

8. The DNA construct according to Claim 7, wherein said regulatory sequence is further capable of directing expression of said DNA in a host cell.

9. The DNA construct according to Claim 6, wherein said DNA is cDNA.

10. An expression cassette comprising:

as components, in the direction of transcription, a first DNA sequence comprising a sufficient portion of the region 5' to the translational start site of a gene expressed in seed to provide for expression of a DNA sequence in seed; a second DNA sequence encoding an oleosin or a sufficient portion thereof to provide for targeting to an oil body, said second DNA sequence including at least one

natural or synthetic restriction site; and a translational and transcriptional termination region; wherein said components are operably linked and expression of said second DNA sequence is regulated by said first DNA sequence.

11. The expression cassette according to Claim 10, wherein said first DNA sequence is from a gene expressed in a cereal or grain seed cell.

12. The expression cassette according to Claim 10, wherein the genome of Aradopsis thaliana comprises at least one of said first DNA sequence and said second DNA sequence.

13. The expression cassette according to Claim 10, wherein said second DNA sequence encodes a peptide from the group consisting of:

(a) a peptide comprising at least eight amino acids of and up to the full sequence of the following amino acid sequence:

```
                    10                  20
M-N-G-R-D-R-D-Q-Y-Q-M-S-G-R-G-S-D-Y-S-K-
                    30                  40
S-R-Q-I-A-K-A-A-T-A-V-T-A-G-G-S-L-L-V-L-
                    50                  60
L-S-L-T-L-V-G-T-V-I-A-L-T-V-A-T-P-L-L-V-
                    70                  80
I-S-S-T-I-L-V-P-A-L-I-T-V-A-L-L-I-T-G-S-
                    90                  100
L-S-S-G-G-F-G-I-A-A-I-T-V-F-S-W-I-Y-K*Y-
                    110                 120
L-L-I-E-H-P-Q-G-S-D-K-L-D-S-A-R-M-K-L-G-
                    130                 140
S-K-A-Q-D-L-K-D-R-A-Q-Y-Y-G-Q-Q-H-T-G-W-
                    150
E-H-D-R-D-R-T-R-G-G-Q-H-T-T; and
```

(b) a peptide that is encoded by a DNA sequence identified by means of an oligonucleotide probe designed based upon said amino acid sequence in (a) or a fragment thereof.

14. An expression cassette comprising:

an oil body protein (OBP) gene which includes a sufficient portion of the region 5′ to the translational start site to provide for expression of said gene in a seed cell and which includes at least one restriction site between just 5′ to the codon for the initiating methionine and 5′ to the translational stop signal of said OBP gene.

15. The expression cassette according to Claim 14, further comprising a DNA sequence encoding a polypeptide of interest inserted into said restriction site in reading frame with said OBP gene.

16. The expression cassette according to Claim 14, wherein said restriction site is a synthetic restriction site.

17. The expression cassette according to Claim 14, further comprising an oligonucleotide adapter coding for a protease recognition site inserted into said restriction site.

18. The expression cassette according to Claim 17, wherein said protease is collagenase.

19. An expression cassette comprising:

a first DNA sequence encoding a polypeptide of interest inserted in reading frame into an oil body protein (OBP) gene which includes a sufficient portion of the regulatory region 5′ to the translational start site of said OBP gene to provide for expression of said gene in seed, wherein said sequence is inserted at a site in said gene so as to be expressed under said regulatory region.

20. The expression cassette according to Claim 19, further comprising a second DNA sequence encoding a protease recognition site 5′ to said first DNA sequence, wherein said second DNA sequence is in reading frame with said first DNA sequence and said OBP gene.

21. A method for obtaining expression of a polypeptide of interest in seed, said method comprising:

transforming a host plant cell with an expression cassette under genomic integration conditions, wherein said expression cassette comprises as components, in the direction of transcription, a first DNA sequence comprising a sufficient portion of the region 5′ to the translational start site of a gene expressed in seed to provide for expression of a DNA sequence in seed; a second DNA sequence encoding an oleosin or a sufficient portion thereof to provide for targeting to an oil body, said second DNA sequence including at least one natural or synthetic restriction site into which is inserted in reading frame a third DNA sequence encoding a polypeptide of interest; and a translational and transcriptional termination region; wherein said components are operably linked and expression of said second DNA sequence is regulated by said first DNA sequence to provide for expression in seed.

22. The method according to Claim 21, wherein the genome of *Thaliana aridopsis*7 comprises at least one of said first and said second DNA sequence.

23. A method for obtaining expression of a polypeptide of interest in seed, said method comprising:

transforming a host plant cell with a DNA construct under genomic integration conditions, wherein said DNA construct comprises a first DNA sequence encoding a polypeptide of interest inserted in reading frame into an oil body protein (OBP) gene which includes a sufficient portion of the regulatory region 5′ to the translational

start site of said OBP gene to provide for expression of said gene in seed, wherein said sequence is inserted at a site in said gene so as to be expressed under said regulatory region, whereby said DNA construct becomes integrated into the genome of said plant cell; and

growing said plant to produce seed whereby said polypeptide of interest is expressed as a fusion protein with the expression product of said OBP gene.

24. The method according to Claim 23, further comprising:

isolating said fusion protein from oil bodies in cells of said seed.

25. The method according to Claim 24, wherein said isolating comprises:

lysing cells of said seed to release said oil bodies; and

disrupting said oil bodies whereby said fusion polypeptide is released.

26. The method according to Claim 25, wherein said isolating further comprises:

contacting said fusion polypeptide with a protease capable of recognizing a protease recognition site in said fusion polypeptide located prior to the N-terminus of said polypeptide of interest.

27. The method according to Claim 26, further comprising:

prior to said contacting, binding said fusion protein to a solid support comprising an antibody capable of binding to the expression product of said OBP gene.

28. A method for obtaining a purified polypeptide of interest, said method comprising:

transforming a host plant cell with a DNA construct under genomic integration conditions, wherein

said DNA construct comprises a first DNA sequence encoding a polypeptide of interest inserted in reading frame into an oil body protein (OBP) gene which includes a sufficient portion of the regulatory region 5' to the translational start site of said OBP gene to provide for expression of said gene in seed, wherein said sequence is inserted at a site in said gene so that expression of said DNA sequence is controlled by said regulatory region, whereby said DNA construct becomes integrated into the genome of said plant cell;

growing said plant to produce seed whereby said polypeptide of interest is expressed as a fusion protein with the expression product of said OBP gene;

isolating oil bodies from the cells of said seed;

disrupting said oil bodies whereby said fusion protein is released; and

purifying said polypeptide of interest.

29. The method according to Claim 28, wherein said polypeptide of interest is other than a polypeptide encoded by a plant genome.

30. The method according to Claim 28, wherein said polypeptide of interest is other than a polypeptide naturally present in an oil body.

31. The method according to Claim 30, wherein said isolating comprises:

collecting an oil-body fraction following lysis of cells from said seed.

32. A plant cell comprising:

an expression cassette according to any one of Claims 10, 14, or 19.

33. A plant comprising cells containing an expression cassette according to any one of Claims 10, 14, or 19.

34. A plant expressing a polypeptide of interest in seed, obtained according to the method of Claim 23.

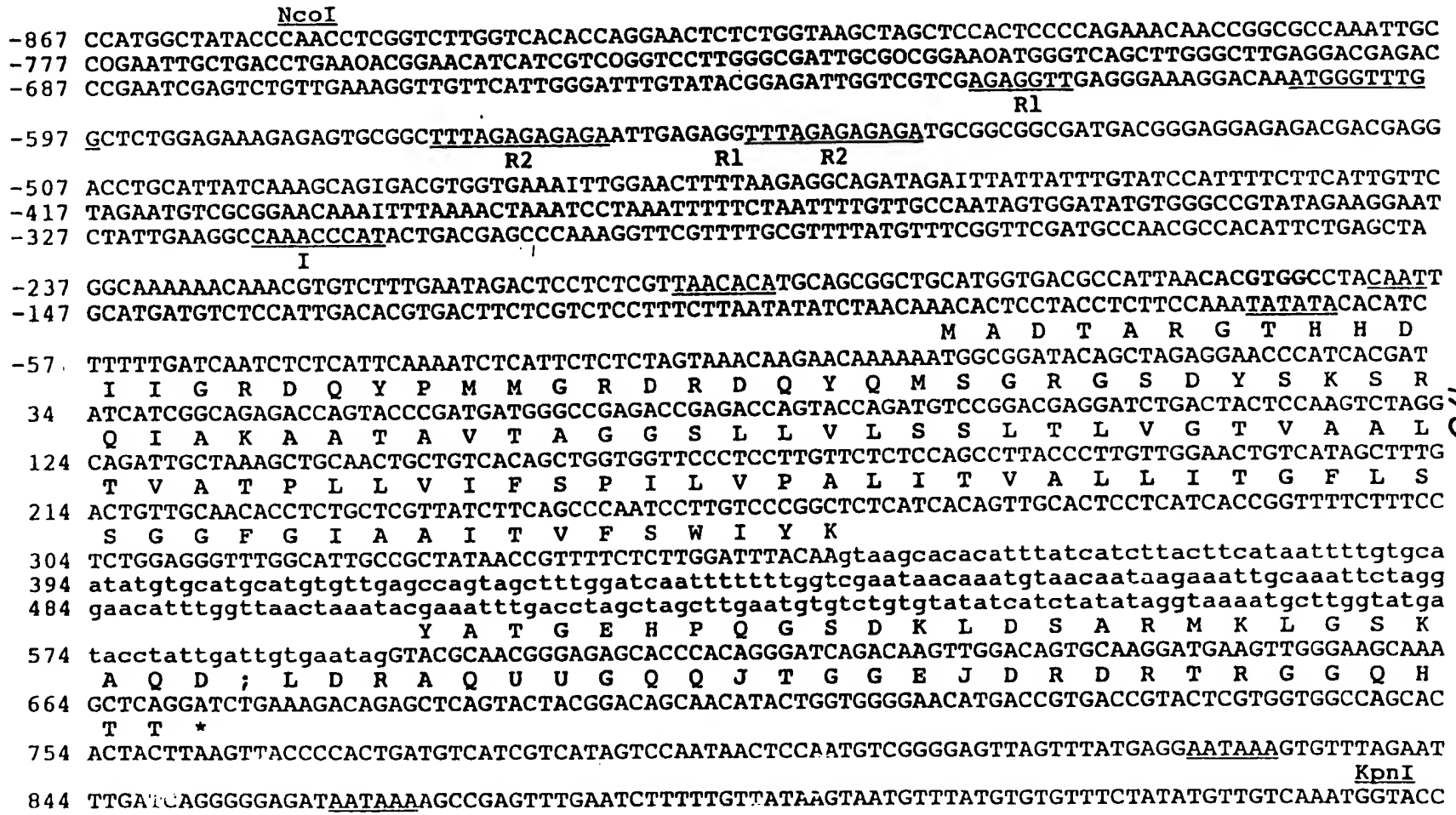
35. Seed comprising an expression cassette according to any one of Claims 10, 14, or 19.

36. Seed expressing a polypeptide of interest, obtained according to the method of Claim 23.

37. A method for obtaining a polypeptide of interest in an oil body, said method comprising:

expressing said polypeptide in seed as a fusion protein with an oleosin, or a sufficient portion thereof to provide targeting to said oil body.

FIG. 1A.



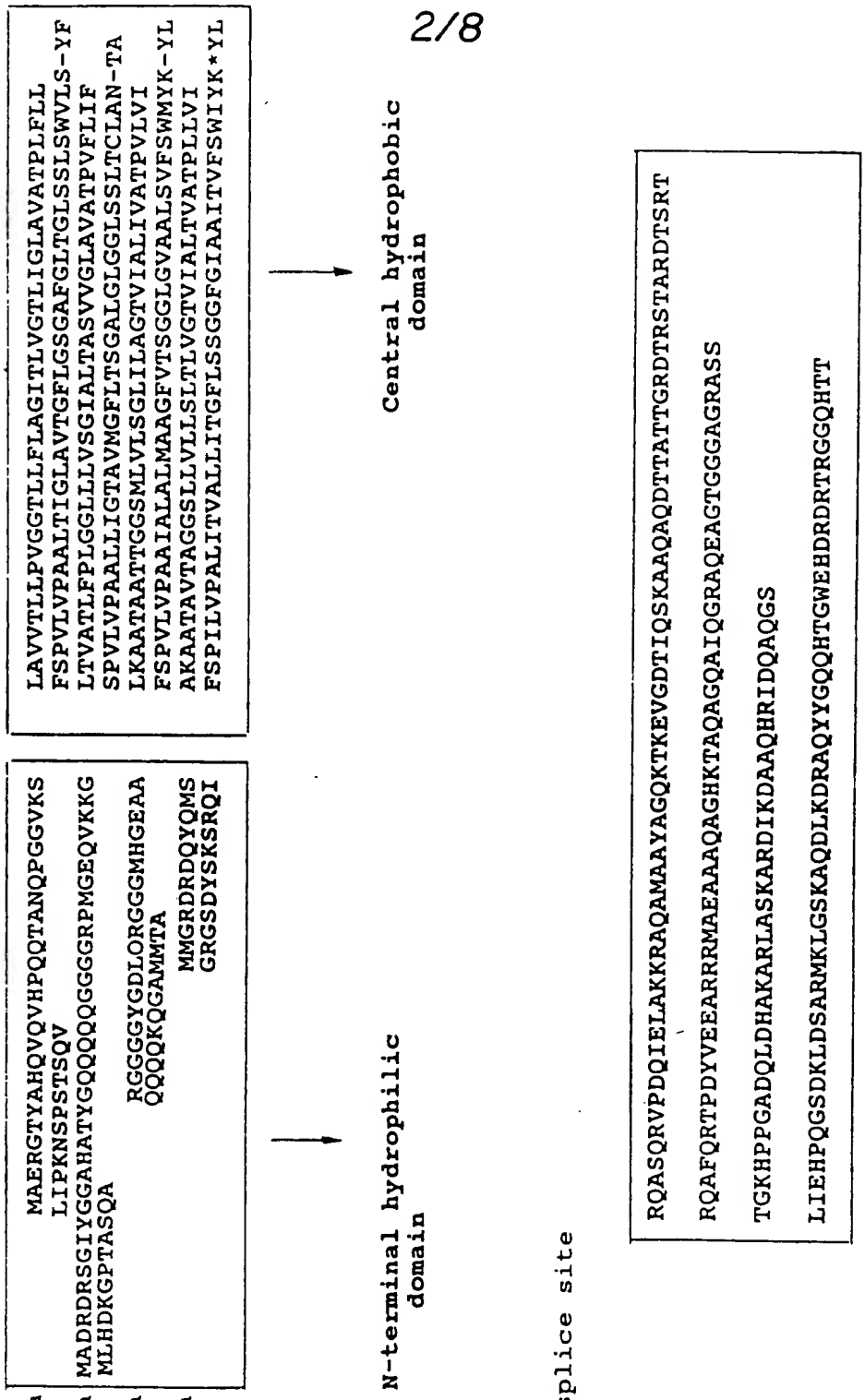
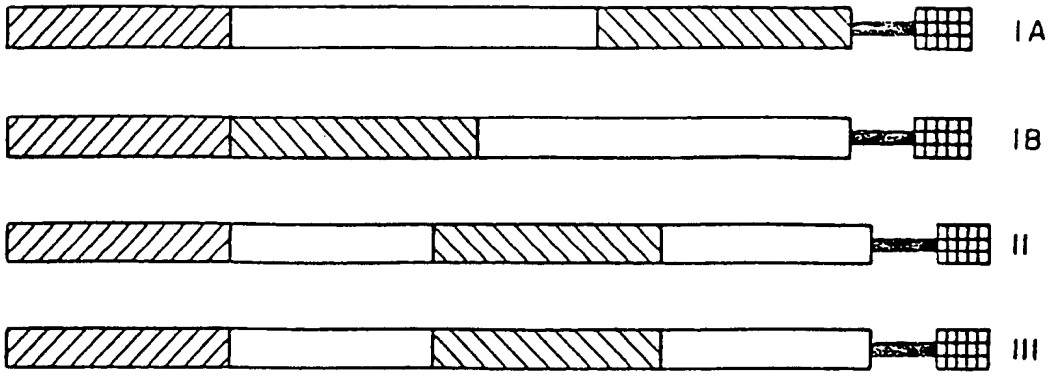


FIG. 1B.



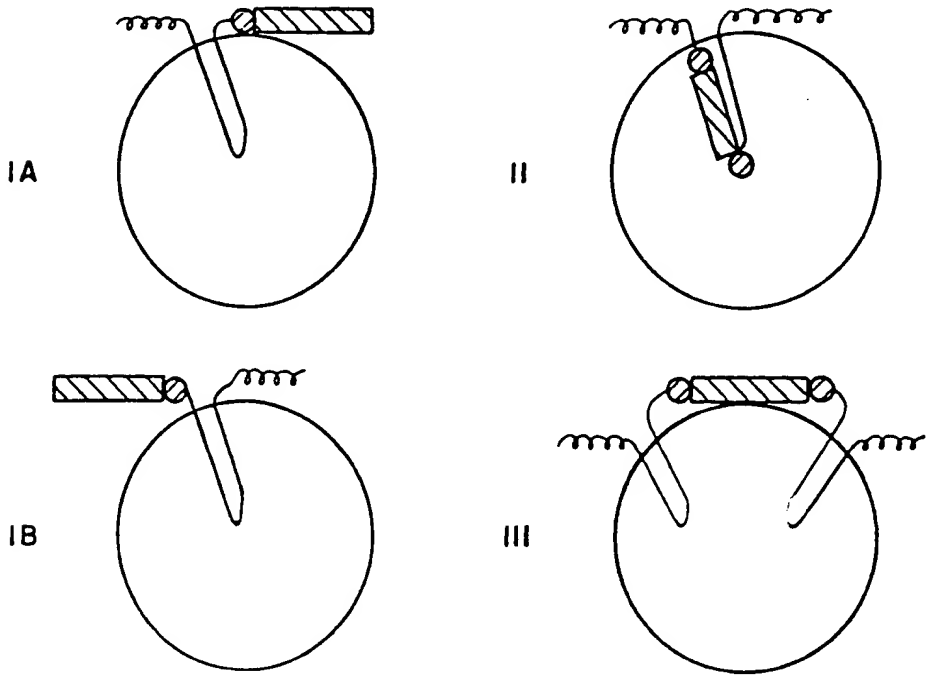


FIG. 2.

```
                Xho I
....AGC ACT GCT CGA GAC ACT TCA AGG ACT....Carrot OBP
                                          (see Hatzopoulos et
    Ser Thr Ala Arg Asp Thr Ser Arg Thr.... al. 1990)

                      Xho I                n times          Nco I
3"..AGC ACT GC                  TCGA  CCG CTC GGT CCG  GC

    TCG TGA CGAGCT                    GGC GAG CCA GGC  CGGTAC

                                      Pro Leu Gly Pro

                                  Collagenase recognition motif
```

FIG. 3.

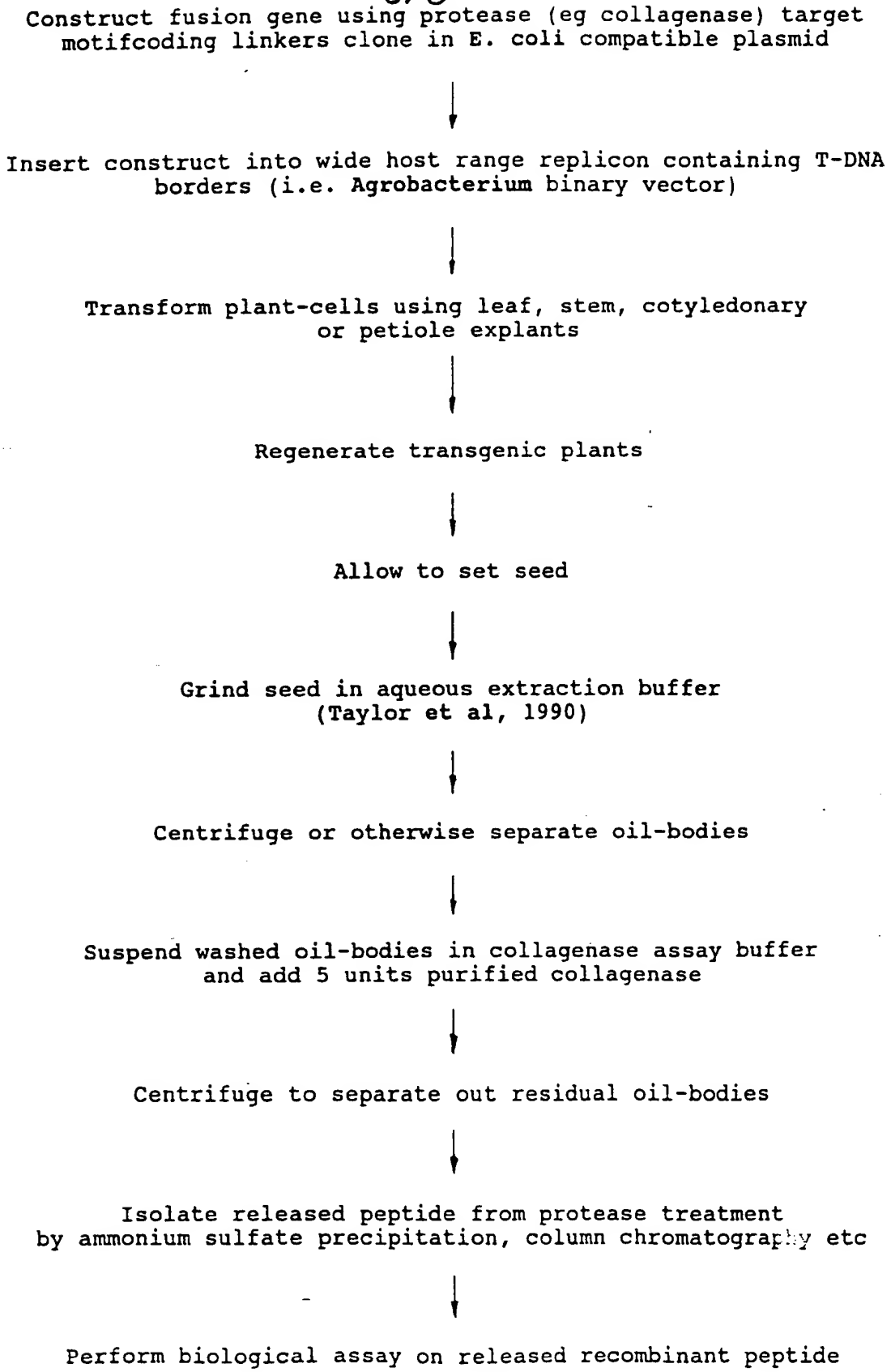


FIG. 4.

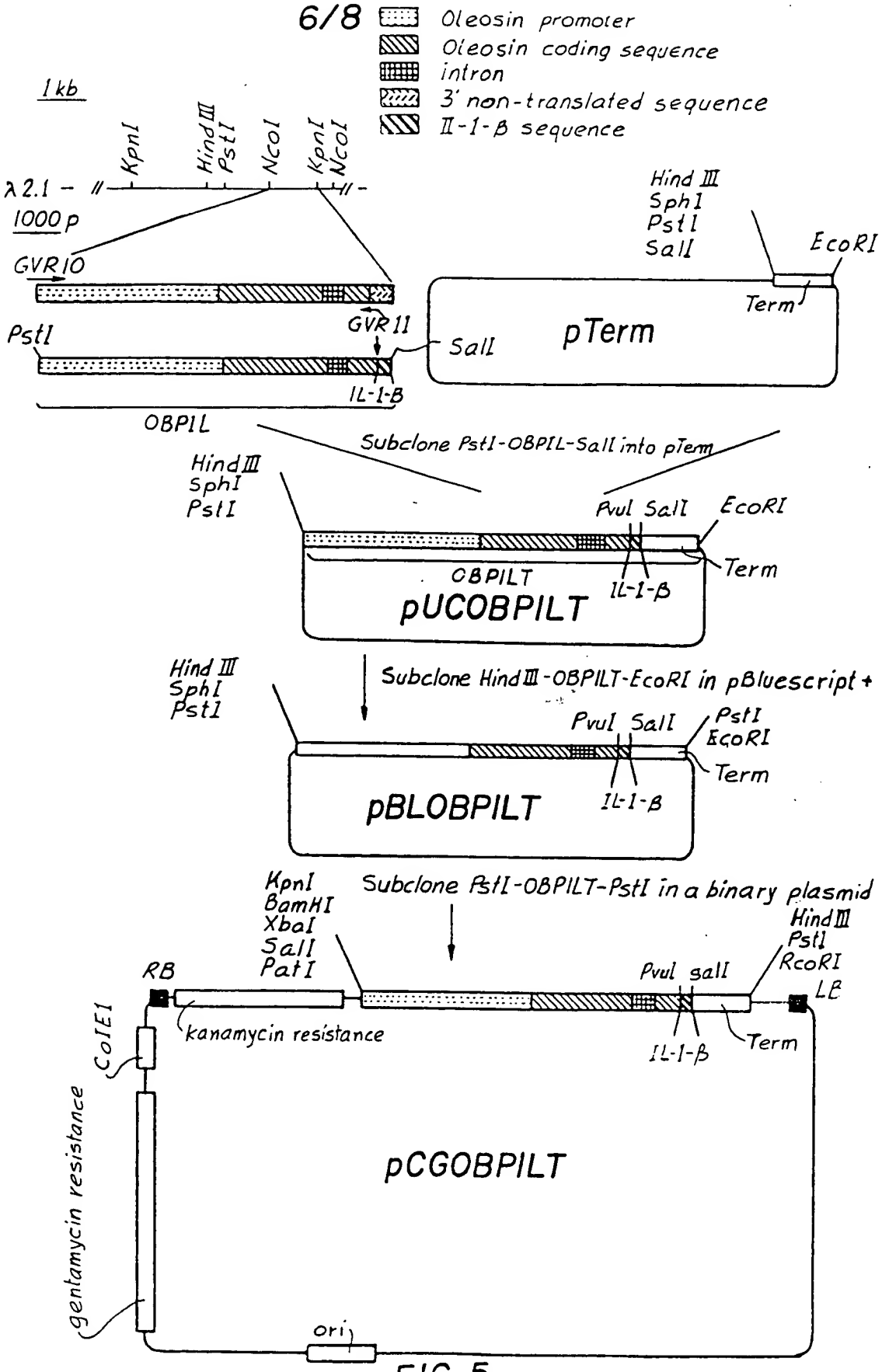


FIG. 5.

SUBSTITUTE SHEET

INTERLEUKIN 1-β gvr11 (71-mer)

a) A. thaliana oleosin Factor Xa

gly gly gln his thr thr ala ile glu gly arg val gln
gly glu glu ser asn asp lys OCH val asp

GGT GGC CAG CAC ACT ACT GCT ATC GAA GGG AGA GTT CAG
PvuI
GGA GAA GAA TCT AAC GAC AAG TAA GTC GAC GG
SalI

b) 3'CCA CCG GTC GTG TGA TGA CGC TAG CTT CCC TCT CAA GT(
[.....
CCT CTT CTT AGA TTG CTG TTC ATT CAG CTG CC 5'
.....]

FIG. 6.

SUBSTITUTE SHEET

PstI

CACTGCAGGAACTCTCTGGTAAGCTAGCTCCACTCCCCAGAAACAACCGGCGCCAAATTGCC
GGAATTGCT
GACCTGAAGACGGAACATCATCGTCGGGTCCTTGGGCGATTGCGGCGGAAGATGGGTCAGCT
TGGGCTTGAG
GACGAGACCCGAATCGAGTCTGTTGAAAGGTTGTTCATTGGGATTGTATACGGAGATTGGTC
GTCGAGAGG
TTTGAGGGAAAGGACAAATGGGTTTGGCTCTGGAGAAAGAGAGTGCGGCTTTAGAGAGAGAA
TTGAGAGGTT
TAGAGAGAGATGCGGCGGCGATGACGGGAGGAGAGACGACGAGGACCTGCATTATCAAAGCA
GTGACGTGGT
GAAATTTGGAACTTTTAAGAGGCAGATAGATTTATTATTTGTATCCATTTTCTTCATTGTTC
TAGAATGTCG
CGGAACAAATTTTAAAACTAAATCCTAAATTTTTCTAATTTTGTTGCCAATAGTGGATATGT
GGGCCGTATA
GAAGGAATCTATTGAAGGCCCAAACCCATACTGACGAGCCCAAAGGTTCGTTTTGCGTTTTA
TGTTTCGGTT
CGATGCCAACGCCACATTCTGAGCTAGGCAAAAAACAAACGTGTCTTTGAATAGACTCCTCT
CGTTAACACA
TGCAGCGGCTGCATGGTGACGCCAT TAACACGTGGCCTACAATTGCATGATGTCTCCATTGA
CACGTGACTT
CTCGTCTCCTTTCTTAATATATCTAACAAACACTCCTACCTCTTCCAAAATATATACACATC
TTTTTGATCA
ATCTCTCATTCAAAATCTCATTCTCTCTAGTAAACAAGAACAAAAAAATGGCGGATACAGCT
AGAGGAACCC
ATCACGATATCATCGGCAGAGACCAGTACCCGATGATGGGCCGAGACCGAGACCAGTACCAG
ATGTCCGGAC
GAGGATCTGACTACTCCAAGTCTAGGCAGATTGCTAAAGCTGCAACTGCTGTCACAGCTGGT
GGTTCCCTCC
TTGTTCTCTCCAGCCTTACCCTTGTTGGAACTGTCATAGCTTTGACTGTTGCAACACCTCTG
CTCGTTATCT
TCAGCCCAATCCTTGTCCCGGCTCTCATCACAGTTGCACTCCTCATCACCGGTTTTCTTTCC
TCTGGAGGGT
TTGGCATTGCCGCTATAACCGTTTTCTCTTGGATTTACAAGTAAGCACACATTTATCATCTT
ACTTCATAAT
TTTGTGCAATATGTGCATGCATGTGTTGAGCCAGTAGCTTTGGATCAATTTTTTTGGTCGAA
TAACAAATGT
AACAATAAGAAATTGCAAATTCTAGGGAACATTTGGTTAACTAAATACGAAATTTGACCTAG
CTAGCTTGAA
TGTGTCTGTGTATATCATCTATATAGGTAAAATGCTTGGTATGATACCTATTGATTGTGAAT
AGGTACGCAA
CGGGAGAGCACCCACAGGGATCAGACAAGTTGGACAGTGCAAGGATGAAGTTGGGAAGCAAA
GCTCAGGATC
TGAAAGACAGAGCTCAGTACTACGGACAGCAACATACTGGTTGGGAACATGACCGTGACC
ACTCGTGGTG
GCCAGCACACTACTGCGATCGAAGGGAGAGTTCAGGGAGAAGAATCTAACGACAAGTAAGTC
GACTCTAG
ACGGATCTCCCgatcgttcaaacatttggcaataaagtttcttaagattgaatcctgttgcc
ggtcttgcga
tgattatcatataatttctgttgaattacgttaagcatgtaataattaacatgtaatgcatg
acgttattta
tgagatgggtttttatgattagagtcccgcaattatacatttaatacgcgatagaaaacaaa
atatagcgcg
caaactaggataaattatcgcgcgcggtgtcatctatgttactagatcGGAATTC

EcoRI

FIG. 7.

SUBSTITUTE SHEET